

The effectiveness of a computer-assisted math learning program¹

Kristof De Witte^{‡62}, Carla Haelermans^{δ‡} and Nicky Rogge^{‡*}

(δ): Maastricht University

TIER, Faculty of Economics and Business
Po Box 6015, 6200 ML Maastricht (the Netherlands)

(‡): Katholieke Universiteit Leuven (KULeuven)
Faculty of Business and Economics

Naamsestraat 69, 3000 Leuven (Belgium);

(*) : Hogeschool-Universiteit Brussel (HUBrussel)
Centre for Economics & Management (CEM)

Stormstraat 2, 1000 Brussels (Belgium)

(¥): IPSE Studies, Delft University of Technology,
PO Box 5015, 2500 GA Delft (the Netherlands)

February 2014

Abstract

Computer-assisted instruction (CAI) programs are considered as a way to improve learning outcomes of students. However, little is known on the schools who implement such programs as well as on the effectiveness of similar ICT-programs. We provide a literature review which pays special attention to the existing causal evidence of computer-assisted programs on learning outcomes. The paper relies on a rich dataset consisting of (i) pupil-level information on the use of a Dutch computer-assisted program and (ii) detailed school-level information on, among others, outcomes on national exams. The results suggest that schools with lower educational attainments use more frequently computer-assisted instruction programs. This suggests that they use CAI-programs to catch-up on learning outcomes. Moreover, using an instrumental variable design, we argue that given the participation in the CAI-program, making more exercises leads to higher test results. Working with a CAI-program seems therefore effective.

Keywords: Computer-assisted instruction; Effectiveness; Mathematics; Secondary education.

JEL-classification: I21; D61

¹ We would like to thank Andre de Moor, Rob Sudmeijer, Irene Pouw, Roelie Kruis, Eline Sneyers and Wim Groot for valuable comments and discussions. Kristof De Witte and Carla Haelermans gratefully acknowledge financial support from the Dutch Ministry of Education, Culture and Science.

² *Corresponding author.* Tel.: 003216326566
E-mail address: Kristof.Dewitte@kuleuven.be; k.dewitte@maastrichtuniversity.nl

1. Introduction

Investment in information and communication technology (ICT) infrastructure was one of the key priorities of education policies of countries all over the world over the past decades. In the EU, for instance, the European Commission developed several action plans which aimed at implementing and integrating ICT into primary schools, secondary schools, colleges and universities. An example of such an action plan was the e-Learning Initiative (European Commission, 2000). This policy was adopted by the European Commission in the year 2000 with the primary purpose of improving the quality of learning by increasing pupil access to ICT facilities in schools. As a result, each European Member State developed national strategies to foster the integration and use of ICT in education. The core objective of most of these strategies was on investing public resources in the implementation of ICT in education. Overall, the results of these policies are positive for the EU Member States. Several academic and non-academic studies have indicated that schools are much better equipped with ICT equipment compared to a decade ago. The ICT infrastructure in schools improved considerably both quantitatively as well as qualitatively. For example for the Netherlands, the ratio of computers to pupils increased gradually to an average figure of one computer for every five pupils (European Schoolnet, 2012). As noted in the country profile report of the Netherlands (European Schoolnet, 2012), the schools in the Netherlands are among the best equipped schools in the EU in terms of ICT infrastructure. Moreover, whereas before there were considerable differences in the availability and the quality of the ICT infrastructure across schools, these differences became much smaller in the Netherlands.

With significant amounts of resources (both public and private resources, yet, for the Netherlands it concerns mostly public resources) being spent on hardware and software in the classrooms of primary and secondary schools, there is an increasing call for accountability on the school administrators, the teachers and the pupils. It is asked whether the large investments on ICT have paid off in terms of improving pupil learning outcomes. Essentially this boils down to answering the question whether the use of technology in schools in general and in classrooms for both teaching and/or learning, more in particular, has actually enabled pupils to realize better learning outcomes. The debate is no longer on whether or not computers should be integrated into the educational system. Rather, the debate is on how the use of educational technology in teaching and learning impacts the pupils' learning outcomes, attitudes and experiences. This shift in focus resulted in an increasing number of empirical research-based studies on the effects of educational technology.

A controversial debate.

The debate on the role of educational technology in the classroom is a long-standing and highly controversial one. One of the key reasons for this continuing controversy seems to be the involvement of multiple stakeholders (such as pupils, pupils' parents, teachers, school management, policy makers, educational experts, etc.) with sometimes diverging interests. Particularly, the question of whether or not the use of educational technology has benefited the pupils' knowledge or the learning experience has stirred a lot of controversy. As in all interesting debates, there are both believers and non-believers.

The believers argue that educational technology when used properly in the classrooms may provide support to the teacher in teaching the course material and help the pupil in mastering the required concepts more easily. They typically refer to studies confirming a positive relationship between the use of educational technology in teaching and learning and the pupils' learning outcomes, their attitudes towards and their experiences with learning. One of the first proponents of the use of technology in the classroom was the psychologist Skinner. In the 1950s, Skinner published several papers (Skinner, 1954, 1958) in which he explains his belief that the use technology in teaching and learning (he uses the term 'teaching machines') might benefit the learning efficiency of pupils. Thanks to the possibilities of repetition and class room differentiation, advocates have invoked that the use of educational technology in teaching and learning may, among other things, help pupils in putting greater focus on understanding the more difficult and complex concepts (Doerr & Zangor, 2000) as well as help them in developing a conceptual understanding of such concepts (Kaput, Hegedus & Lesh, 2007; Kebritchi, Hirumi & Bai, 2010). Other points of strength of educational technology are the interactive nature with high interaction frequency between teacher-system-pupil and the adaptive nature which enables to customize instruction and feedback for the needs of the individual pupil (Wenglinsky, 1998; Woolf, 2009; Shute & Zapata-Rivera, 2007). Wenglinsky (1998), for instance, discussed that educational software can provide pupils with the opportunity to self-organize their learning. Most proponents are also convinced that educational technology can play an important role in the democratization of access to education in the sense that it can enable pupils in different settings (particularly students in more disadvantaged settings) to have more and better learning opportunities (Kaput, 1997). These positive effects in mind, they believe that investing considerable (public or private) resources in the implementation of educational technology in schools is worthwhile.

The non-believers and the critics (e.g., Fuchs & Woessmann, 2004; Honey et al., 2000) are much less enthusiastic and more skeptical about the use of technology in the classroom. They (strongly) contest this alleged positive impact of educational technology on the pupils' achievements and warn that the use of educational technology should not be considered a panacea to the problem of improving pupils' learning outcomes in education. They believe that if there is an impact on learning effectiveness then at best this impact is only marginal. Some of the skeptics even believe that this idea of a positive impact of educational technology on the teaching and learning efforts of teachers and pupils is just a notion invoked by certain stakeholders who benefit from the presence of this perceived link (not the least the developers of such educational hardware and/or software). Some of the non-believers even strongly argue against the use of technology in classrooms thereby claiming that the pupils' test achievements may even be negatively associated with technology use. They fear among other things that the use of computers or other educational tools in the classroom may distract pupils instead of helping them in mastering educational concepts. Another point of worry is that the use of educational technology may undermine the teacher-pupil relationship and reduce the interaction between teachers and pupils.

Evidence.

Summarizing the findings of earlier literature is an intricate matter. Overall, it seems that the evidence is inconclusive with some studies indicating positive effects (e.g., Kulik, 2003; Murphy et al., 2001), other studies showing no strong impact (e.g., Angrist & Lavy, 2002) and some studies finding negative effects (e.g., Spiezia, 2010; Campuzano et al., 2009) of using computer software in teaching and learning. The mixed results are certainly to some extent due to the complexity of the relationship between ICT and learning. Other reasons are the wide variety of assumptions that have been made by research studies and the fact that the impact of educational technology has been studied from different perspectives (e.g., pedagogical, sociological, computer sciences and economics), in different teaching and learning environments and using different methodologies. All of this makes that the findings of one study about the effectiveness of educational technology cannot be generalized beyond the teaching and learning context in which the study was performed. Moreover, as remarked by some researchers (e.g., Cheung and Slavin, 2013; Becta, 2007; Cox and Marshall, 2007), a large majority of the past studies suffer from design flaws and methodological or conceptual weaknesses which raise doubt about the validity of their findings. Several of these limitations are discussed and tackled in this paper.

Contributions.

The current paper contributes to this expanding literature in several ways. First, the primary objective of this paper is to examine the effectiveness of an online and adaptive educational tool in learning mathematics in secondary school. The key feature of this computer-assisted instruction (CAI) tool is that it provides each student with an individual training package based on his or her test results. This training package consists of a wide range of explanatory movies (e.g., screencasts), theory and exercises. We exploit the rich data set, which is logged by the program, to examine the influence of CAI-tool on student test scores.

Second, we address some of the methodological concerns typically observed in past research studies. Using an instrumental variable technique, we focus on causal evidence. The instrument is deduced from the way teachers deal with the CAI-tool. We exploit the fact that some students are more exposed to the program than other students. Using instrumental variables with class and school fixed effects, we obtain causal evidence on the relationship between making online exercises and test results.

In particular, this paper tests two research questions:

1. Do schools with lower educational attainments use computer-assisted instruction programs more frequently?
2. Does more intense exposure to the computer-assisted instruction program cause higher test scores?

This article is organized as follows. The ensuing section presents the main findings of previous studies on the association between ICT use and pupils learning outcomes. We focus on both the different methodologies as well as on literature on ICT in general and CAI-tools in particular. Section

3 presents the data and introduces the computer-assisted tool. Section 4 presents the main results for the first research question, while section 5 presents the results for the second question. The final section concludes by summarizing the key findings of this article and providing policy recommendations.

2. Literature review

There are several dimensions along which the literature on the effectiveness of computer-assisted instruction (CAI) can be classified. In this review, we focus on the effectiveness of CAI-tools for mathematics.

Quantitative versus Qualitative studies.

Qualitative studies frequently use semi-structured, in-depth interviews to collect information about the perceptions, attitudes, or opinions of the different stakeholders in education (e.g., the pupils, the teachers, the school directors, educational experts, etc.). Examples are Schacter (2001), Sivin-Kachala (1998), and Reimer and Mayer (2005). Reimer and Mayer (2005), for instance, employed a qualitative approach to investigate the impact of computer-assisted instruction (CAI)-tools in the teaching and learning of mathematics. The qualitative analysis consisted of interviewing pupils as well as administering an attitude survey among them to examine the impact of CAI-tools. In particular, the effect of virtual manipulative computer applets on 3rd grade pupils' achievement levels and attitudes. The results of this qualitative examination showed that the use of computer-based virtual manipulation in teaching mathematical concepts to pupils helps pupils in understanding fractions. One of the explanations for this result, as discussed by Reimer and Mayer is that the use of computer-based virtual manipulation enables the teachers to provide the pupils with more immediate and individual-specific, and hence better, feedback.

The majority of the studies in the literature used a quantitative analysis approach to examine the impact of technology in education. Typically, studies employed statistical analysis techniques such as simple correlation analysis (McAlister, Dunn and Quinn, 2005), regression techniques (Angrist and Lavy, 2002), (M)AN(C)OVA (Pilli and Aksu, 2013), and randomized control trial designs (Potocki, Ecalle, and Magnan, 2013; Papastergiou, 2009). The results of the qualitative studies are summarized in multiple interesting meta-analyses. Examples of meta-analyses include Kulik and Kulik (1991), Kulik (1994, 2003), Murphy et al. (2001), Blok et al. (2002), Christmann and Badget (2000, 2003), Goldberg, Russell, and Cook (2003), Rayne and Baggott (2004), and Cox and Abbott (2004). We briefly describe the main findings of the most recent meta-analyses (i.e., the meta-analyses that appeared since 2000).

Murphy et al. (2001) considered 195 (quasi-)experimental studies conducted in the nineties. Minimum methodological requirements were imposed to select among these 195 studies the ones qualified for more detailed analysis. This resulted in a subset of 31 studies. Based on the outcomes of these studies, Murphy et al. (2001) computed an impressive average effect size of 0.45 for mathematics (i.e., as a proportion of the standard deviation of the mathematics test scores).

Christmann and Badget (2000) examined the difference in achievement levels between pupils who were taught by the traditional instruction approach (control group) and pupils who had classes in which a CAI tool was used as a supplement to the traditional classes (the experimental group). In doing so, they compiled data from 26 studies. The overall results suggested a mean effect size of 0.127. Hence, pupils who were taught and who learned mathematics via an educational software as supplement to traditional teaching displayed higher achievement levels compared to the other pupils in the control group. Moreover, as denoted by Christmann and Badget, the achievement level of the typical pupil in the experimental group increased from the 50th percentile to the 55th percentile.

Rayne and Baggott (2004) performed a meta-analysis of 40 studies that examined the differences in effectiveness of a 100% traditional teaching approach and a teaching approach which supplements traditional teaching with CAI (hence, a mixture of traditional and computer-assisted teaching). They concluded that the combined traditional-CAI teaching approach was more effective in that it enables pupils to realize higher levels of achievement compared to the pupils who were taught by the 100% traditional teaching approach.

Causal evidence on educational technology.

Typically it is very difficult to estimate causal relations between the use of educational technology and the changes in the pupils' learning outcomes. As nicely formulated by Biagi and Loi (2013, p. 29): "in practice, we seldom have the chance to go beyond measures of association because, even if we have a clear view on the causal relationship between the left-hand and the right-hand side variables, we are not able to identify it through lack of data." We observe in earlier literature various studies focussing on associations (e.g., Fuchs and Wößmann, 2005; Notten and Kraaykamp, 2009; Luu and Freeman, 2011; Kubiato and Vlckova, 2010; Wittwer and Senkbeil, 2008; Spiezia, 2010). However, recently, this trend is somewhat changing. Probably due to more data availability and increased attention to causality, an increasing number of studies employed an experimental or quasi-experimental approach to examine how the use of technology in the classroom, in the school, or at home relates to the pupils' learning outcomes. While correlational studies have their merits, we focus on the causal evidence and the way it has been revealed.

Machin et al. (2007) used data on the educational outcomes in the UK primary schools for the period 1999-2003 to estimate the causal impact of ICT investments made during these periods. To control for the potential issue of endogeneity in the investment variable, they employed an instrumental variable (IV) approach. The IV-analysis revealed a significant positive causal impact. Banerjee, Cole, Duflo and Linden (2007) scrutinized the results of a randomized policy implemented by two regions in India with the objective of improving the quality of education in urban slums. The key finding was that the use of educational technology developed to enforce the mathematical skills of pupils did in fact succeed in realizing this objective. However, the researchers underlined that this positive result was limited to the domain of mathematics. No positive results were found for the pupils' performances in other domains. Leuven et al. (2004) investigated how a subsidy established by the

government in the Netherlands for the purchase of educational technology (both hardware and software) influenced the learning outcomes of disadvantaged pupils in primary school. To examine the impact of this subsidy, the authors exploited the discontinuity in this subsidy (with some of the schools with disadvantaged pupils being eligible for the subsidy and other schools not) to estimate the effect of educational technology on the learning outcomes of disadvantaged pupils. More precisely, using a difference-in-differences framework, they compared the change in pupil performances (i.e., the difference in pre- and post-test scores) between the disadvantaged pupils in schools who received the subsidy and their counterparts in schools who did not. The results showed that the subsidy had a negative impact on the pupils' learning outcomes. In other words, disadvantaged pupils in schools who received the subsidy achieved lower changes in test scores compared to pupils in schools who didn't qualify for this subsidy. Finally, Rouse et al. (2004) focused on the influence of an instruction technology on the reading and language skills of pupils in the US. Using a randomization framework, they found a limited positive impact on the language skills. Dynarski et al. (2007) employed an experimental design in which the changes in pre- and post-test scores are compared between pupils who used various software tools in the classroom (treatment group) and the pupils who did not (control group). In the study, 439 volunteer teachers participated in the experiment. This resulted in data for approximately 9.500 pupils. A comparison of the pupils' pre- and post-test scores between the pupils in the treatment and control group showed that on average there is no considerable difference between the users and non-users. This suggests that the impact of the use of educational technology on pupils' learning outcomes in mathematics and reading is questionable.

Causal evidence for computer-assisted tools.

As remarked by, among others, Beal, Arroyo, Cohen and Woolf (2010), there is a large variety of educational technology, ranging from simple and static ('old style') tools and the more innovative, dynamic, interactive, and flexible tools. The more simplistic tools are typically less flexible and less interactive. They aid pupils in certain standard tasks such as performing computations (examples are calculators, excel-software, etc.). The most recent educational technology is more flexible and more interactive in the sense that it is adaptive to the needs of each individual pupil. Examples of such instruments (hardware and/or software) include *LOGO*, *Derive*, *Cabri*, *Mathematica*, *Coypu*, *Geometric Supposer*, *Geometer's Sketchpad*, *Cognitive Tutor Algebra*, *Larson Algebra*, and *Plato Algebra*, *Frizbi Mathematics 4*, and many others. The effectiveness of most of these software tools has been examined by research studies in (quasi-)experimental designs.

Dynarski et al. (2007), for instance, investigated by an experimental design how the use of *Achieve Now*, *iLearn Math*, and *Larson Pre-Algebra* in the teaching of mathematics in the sixth grade was related to the pupils test outcomes. The three software products were developed primarily for providing tutorial and practice opportunities. Overall, the results suggested that there were no significant differences in the test scores between users and non-users. Nevertheless, large differences were observed across schools. In addition, a series of statistical tests indicated that the

included classroom-, teacher- (e.g., teacher experience, teacher gender, teacher education level, etc.), and school-level characteristics were not statistically significantly related to the observed differences in test scores.

Pilli and Aksu (2013) employed a quasi-experimental research design to examine the impact of educational software for mathematics on 4th grade pupils' achievements in mathematics, the pupils' attitudes towards mathematics and computer-assisted teaching and learning as well as the retention of mathematical knowledge. They found that the educational software is an effective tool for teaching and learning mathematics in the sense that pupils who used the software in the classrooms achieved higher test scores and had more positive attitudes towards mathematics.

Roschelle et al. (2010) focus in an experimental design on a software tool that was developed with the purpose of enabling a large group of pupils to learn more advanced mathematical concepts and skills in Texas. They identified a positive significant impact of the use of the program on pupils mathematics achievements. Roschelle et al. conclude that the CAI-tool is an effective tool to enhance pupil knowledge of more advanced mathematics.

Edwards and Quesada (2007) argued that *Cabri3D* offers, among other things, three important advantages in the teaching and learning of mathematics. One such an advantage is that the visualization aspects of the software tool help pupils in better understanding three-dimensional figures and shapes. Another advantage is that it provides information which helps students in understanding the relationship between two- and three-dimensional concepts. The effectiveness of the regular version of *Cabri* as educational software in teaching and learning of mathematics was examined more recently by Köklü and Topçu (2012). They focused on the impact of *Cabri* among 10th graders and found that pupils who used *Cabri* had a better understanding of the concepts about graphs of quadratic functions. More precisely, whereas pupils who were taught these concepts by the traditional approach had more difficulties in understanding these concepts, the ones who used *Cabri* in their learning experienced fewer difficulties.

Koedinger, McClaughlin, and Heffernan (2010) investigated the effect of *ASSISTments* on the math test scores of seventh grade pupils in middle school. *ASSISTments* is a web-based tutor system designed for teaching mathematics. A key feature of this system is that it aims at addressing the need for timely pupil assessment while at the same time providing instruction to the pupils. By doing so, the tutor system avoids the loss of instruction time that typically occurs during assessment. Koedinger et al. (2010) used a quasi-experimental approach to measure the effect of the *ASSISTments* tool. The sample consisted of 1,240 pupils. Koedinger and his colleagues found that the use of the web-based tutor system for teaching mathematics is effective in improving pupils' learning of mathematics. Moreover, the comparison between the improvements in math test scores of pupils in the treatment and control group also indicated that the largest improvements in the treatment were obtained for special education pupils. This suggests that the *ASSISTments* system is particularly effective for this group of pupils as it enables them to catch-up (at least to some extent) with the other pupils. Note that a similar result was also found by Bouck and Flanagan (2009) for the use of other types of assistive technology in the teaching and learning of mathematics.

3. Computer-assisted tool and Data

The computer-assisted tool.

This paper considers a Dutch computer-assisted tool called *Gotit?!*. The hallmark of the education software *Gotit?!* is that it was created through consideration of the best approaches of teaching mathematics as well as the needs (cognitive, psychological, etc.) of the students. The CAI-tool offers a large amount of exercises of different difficulty levels. This enables each pupil to organize the work and progress at a rate consistent with his/her own level of ability. This allows the teacher to differentiate within the class. Pupils who experience fewer difficulties with the theory and advance quickly in solving exercises can go to exercises of higher difficulty level without being slowed down by pupils who progress more slowly.

The CAI-tool is adaptive in that it adjusts its exercises to the knowledge and level of the student. *Gotit?!* provides pupils with tips on organization and skills for solving exercises. All of this may benefit the pupil's confidence in the learning content, improve their meta-cognitive skills and provide a way for skill-drill (i.e., practicing an activity until it becomes automatic). The content is organized along 11 subjects. These include, e.g., additions, multiplications or counting principles.

On top of this, the *Gotit?!* system offers features which give quick and continuous feedback to the teacher on pupil learning progress both at the level of the individual pupil as well as the classroom. More precisely, the feedback and control system comprises tools for tracking each individual pupil's step-by-step progress so that at each moment an accurate identification of his/her competence level is possible. In this way, the teacher can monitor which pupils realize the milestones and which pupils require additional attention. Based on this continuous stream of information on pupil progress, the teacher can determine whether an adjustment in the instruction approach or any other type of remediation is warranted for the class as a whole or for one or more individual pupils. In addition, *Gotit?!* also includes communication features which enable teachers to interact and communicate with the pupils both at classroom level as well as individually. Depending on the circumstances, the teacher can decide to provide feedback to all pupils in the class, a subgroup of pupils or just one individual pupil.

The sample.

Various studies in earlier literature suffer from the use of small samples. Numerous studies, for instance, used sample sizes of less than 50 pupils. Researchers typically experience a dilemma in which they have to trade off the choice for a large-scale study with the choice for a detailed study of the impact of very specific uses or types of educational technology (i.e., a particular educational software product). This paper does not suffer from this drawback. Data were provided by the publisher that developed the *Gotit?!* Software (i.e., ThiemeMeulenhoff). After removing data with incomplete cells or incorrect logs of time (we removed students whose recorded time to complete the pretest was more than 3 hours, while the median student took less than 1 hour for the test), the

cross-sectional sample consists of 9,898 pupils in the first three grades of secondary education in 2012. In addition, we augmented the pupil-level data with school-level information from the Ministry of Education, Culture and Science. The most recent year available is school year 2011-12. This information provides us with additional insights in the educational attainments of the school, the allocation of the school budget and the composition of the school in terms of share of students from disadvantageous backgrounds. Even more importantly, this data source provides us with information on the school average of the national and school exam. In the final years of secondary education, all students in the Netherlands have to take two exams for each course in which they received lessons (independent of the educational track). The former exam – the ‘national exam’ – is an absolute assessment with criterion- referencing which is uniform for all subjects and schools in the Netherlands (see De Witte, Geys and Solondz, 2013 for a discussion). The latter exam – the ‘school exam’ – has fewer quality controls in its construction and evaluation as it is set up and corrected only by a school’s teachers. Aggregate information on the school and national exam is publically available.

Descriptive statistics.

The descriptive statistics are presented in Table 1. The ‘average pre-test’ is computed as the average of the pre-tests of all students at a school location. It ranges between 28 and 73, with an average of 59. The ‘post-test scores’ consist of the average of the various subjects the student took. It ranges between 0 and 1. The ‘number of exercises’ the student took in the CAI-tool vary between 1 and 1248. Despite this high maximum, 75% of the students took 62 exercises.

We further have information on the student’s ability. The pre-test that students took ranges between 4 and 96, with an average of 59. Students took this pre-test in, on average, 55 minutes. 25% of the student took maximum 38 minutes for the pre-test.

We observe a broad set of control variables at the school level. First, the school average of the national exam amounts, on average, to 6.35. This is slightly lower than the mean of the school average of the school exam, which equals 6.45. It is commonly observed in the Netherlands that the school exam is slightly higher graded than the standardized national exam. The number of teachers (expressed in full time equivalents) is on average 160 per school (note, this denotes the school group rather than the school location). This should be compared to the average number of students per school, which amounts to 2467. This indicates that the average class size counts about 15 students. The dropout percentage is standardized such that the median school has a percentage of 1. We observe information on the costs for materials (expressed in million euros), the percentage of students coming from disadvantaged neighborhoods (mean 5.1%) and the percentage of students in supportive ability tracks (mean 12%).

Finally, we observe in the data 128 school groups with 171 school locations, 1,947 different classes and 2,239 different combinations of school locations and classes.

Table 1: Descriptive statistics

Variable	N. obs.	Mean	S.D.	Min	.25	Median	.75	Max
Variables of interest								
Average pre-test (school level)	9898	59.17	4.94	28	56.88	59.25	62.2	73
Post-test scores (student level)	9898	0.64	0.19	0	0.53	0.66	0.8	1
Number of exercises	9898	50.52	65.53	1	14	30	62	1248
Student ability								
Pre-test – score (student level)	9898	59.11	12.32	4	52	60	67	96
Pre-test - time in seconds (student level)	9898	3320.85	2034.67	0	2280	3409.5	4529	9996
Instrument								
Exposure to CAI - Continuous	9898	3.54	2.82	1	1	3	5	11
Exposure to CAI - Dummy (little - extensive)	5478	0.1	0.3	0	0	0	0	1
Control variables								
Average national exam (school level)	9898	6.35	0.17	5.78	6.28	6.37	6.44	6.71
Average school exam (school level)	9898	6.45	0.12	6.08	6.38	6.46	6.53	6.8
School size (school level)	9898	2467.94	1169.59	400	1563.00	2284.00	3249.00	5641.00
Number of teachers per student (1000 fte - school level)	9898	0.16	0.08	0.02	0.11	0.15	0.21	0.38
Costs for materials (m euro - school level)	9898	2.35	1.62	0.35	1.18	1.91	2.79	7.91
Dropout percentage (school level)	9898	1.05	0.46	0.17	0.72	0.99	1.31	4.6
% disadvantaged students (school)	9898	5.14	7.66	0	0.1	1.4	6.83	41.8
% supportive ability track (school)	9898	11.99	11.58	0	0	9.42	21.33	45.67
Fixed effects								
Class	9898			1				1947
Class and school	9898			1				2239

4. Do schools use computer-assisted learning tools to improve learning outcomes?

Model specification.

To examine the first research question, we correlate average school attainments to the intensity of use of the CAI-tool.

$$\text{Intensity_use}_i = \text{cte} + \beta_1 \text{School_attainments}_j + \beta_2 X_{i,j} + \eta_j + \varepsilon_i \quad (1)$$

Where i , j and k denote the student, school location and class level, respectively. The dependent variable is measured at the student level. We proxy the intensity of working with the CAI-tool in two ways. First, we consider the number of exercises a student has made. The more exercises, the more intense the student has worked with the tool. Second, we consider the number of subjects the student has successfully completed. By combining the two outcome variables complementary information is obtained. The school attainments, as the independent variable, is measured by the outcomes on the school exam and the nation-wide standardized national exam.

In Equation (1) we further control for observed heterogeneity X_{ij} . This includes, first, the student attainments on the pre-test. Including pre-test information is important to capture potential endogeneity arising from unobserved student ability. More able students require less exercises. We observe a significant, though low, correlation between the pre-test (student level) and the nationwide exam (school level). In a similar vein, we include as a second control variable the time the student needs to write the pre-test. Whereas the first two control variables are at the student level, the other control variables are at the school (location) level. They include the level of early school leaving, the percentage of students from disadvantaged neighborhoods (APCG), and the percentage of students at the school in supportive ability tracks (LWOO).

Finally, we include class fixed effects η_j and an error term ε_i in the regressions. Earlier literature argued that including the fixed effects is important as it captures the nested structure in the data. Pupils are being nested within classes (and, hence, teachers) and schools. As discussed by Roschelle et al. (2010) not accounting for this nesting can be an important limitation as it does not rule out the presence of clustering effects in the results.

Results.

The results are presented in Table 2. The first two model specifications provide the results for the number of completed subjects as outcome variable. The last two model specifications have the number of exercises as outcome. We observe that the higher the average national exam grade, the less subjects and exercises are completed. The same yields for the school exam. On the opposite, students in schools with lower national exam outcomes make more exercises and more subjects – controlled for individual abilities and class fixed effects.

We further observe that, at the individual level, higher pre-test scores are positively correlated with the intensity that the program is used. This is no longer significantly different from zero if the number of exercises is used as an outcome. In addition, the faster the student worked in the pre-test, the less subjects and exercises he/she completed. It should be noted that all the estimations include class fixed effects such that observed and unobserved heterogeneity at class level (e.g., due to the teacher, peer-effects, or ability tracking) is accounted for.

Additional variables are added to Models 2 and 4. This confirms the earlier results. In addition, it shows that the more disadvantaged students a school has, the more intensive the CAI-tool is used. It is also remarkable that in schools with more teachers per student, there are less subjects and exercises made. While this correlation is insignificant, it might weakly suggest that teachers and CAI-tools are substitutes.

The findings confirm the first hypothesis: schools with lower educational attainments use computer-assisted learning programs more frequently, in order to catch-up in learning outcomes.

Table 2: Relationship between learning outcomes and intensity of program use

	Model 1	Model 2	Model 3	Model 4
Independent variable	subject	subject	exercises	exercises
Average national exam (school level)	-4.4697***	-0.0136	-89.6699***	-37.611
Average school exam (school level)		-3.751***		-43.9077
Pre-test – score (student level)	0.0407***	0.0391***	0.0933	0.0871
Pre-test - time (student level)	-0.0002***	-0.0003***	-0.0047***	-0.0048***
Average pre-test (school level)		0.0566**		0.14
Number of teachers per pupil (fte - school level)		-0.288		-106.419
Costs for materials per pupil (school level)		-0.016		-180.326
Dropout percentage (school level)		-0.357		-4.649
% disadvantaged students (school)		0.0367***		0.736**
% supportive ability track (school)		-0.0135		0.0592
Constant	30.3286***	23.647***	629.9820***	565.3156**
Fixed Effects	Class	Class	Class	Class
Number of observations	9898	9898	9898	9898
R ² -adjusted	0.1447	0.1376	0.176	0.174

legend: * p<0.05; ** p<0.01; *** p<0.001

5. Does more intense exposure to the CAI-tool cause higher test scores?

Instrumental Variable Analysis.

The intensity by which a student participates in the CAI-tool is in an unobserved way correlated to the extent to which the school, and in particular the teacher, stimulates the use of the software. This is also acknowledged by Hennessy, Ruthven and Brindley (2005) who scrutinized the role of the teacher and found some evidence that the teacher plays an important role in the way that educational technology is used in the classroom. More specifically, the teacher's attitude towards, as well as his acquaintance with, the use of educational technology determines to a considerable extent (1) what educational technology is chosen, (2) how the educational technology will be used in the classroom, and (3) how the pupils will use the technology. This finding of the teachers' crucial role in the implementation and the choice of use of educational technology has consistently been found across earlier studies (see Section 2). Given the unobserved heterogeneity, simply regressing the number of exercises by the student on the test scores, would therefore be an endogenous regression. This means that the:

$$\text{Cov}(X, \varepsilon) \neq 0$$

Due to the lack of experimental data, an instrumental variable approach is the best procedure to remove the endogenous part in the regression. To obtain an instrument, we start from earlier work by Spiezia (2010) and Rouse et al. (2004). The former was able to make a good distinction between the different uses and the intensity levels of usage. The latter apply the participation in an CAI-tool as an instrument. In the current setting, this implies that the intensity of the participation in the

program would be a good instrument (denoted by Z). We make this operational by the number of passed subjects. For example, a student who wrote only two subjects is less exposed to the program than a student who wrote subject tests for all subjects. The number of subjects that the student passed is therefore strongly correlated to the number of exercises in the CAI-program (significant correlation at 1%-level of .36), while it does not have a direct effect (although an indirect via the number of exercises) on the test scores. It therefore fulfills two conditions of an instrument:

Exogeneity condition: $\text{Cov}(Z, \epsilon) = 0$

Relevance: $\text{Cov}(X, Z) \neq 0$

We define the instrument in two complementary ways. Both specifications are at the individual level. First, we consider the participation in the program as a dummy variable. Only students participating to all or all-but-one of the subjects (instrument = 1) and students participating to less than 2 subjects (instrument = 0) are included in the analysis (resulting in 5478 observations; see Table 1). A second specification of the instrument consists of considering the number of passed subjects as a continuous variable. From the descriptive statistics in Table 1 we learn that the average student wrote three subjects, while 75% of the students passed 3 subjects.

First stage tests.

Although the exogeneity condition cannot be directly tested because it involves a correlation between the instrument and an unobserved error, we can follow some standard tests to examine its adequacy.

First consider the (Durbin-)Wu-Hausman test (numerically equivalent to the standard Hausman-test) which examine whether the OLS and IV estimates are different. If they differ significantly, we can conclude that X is an endogenous variable. In our application, the regressor is clearly endogenous as the Wu-Hausman F test equals for the dichotomous specification 130.32 [$F(1,6547)$] and 462.92 for the continuous specification [$F(1,12022)$]. The Durbin-Wu-Hausman Chi-squared test equals 127.86 [$\text{Chi-sq}(1)$] and 445.91, respectively.

Second, the Anderson LM-statistic, which tests the underidentification, equals 3718.93 for the dual instrument and 5914.80 for the continuous instrument, such that the equation is identified, i.e., the excluded instruments are correlated with the endogenous regressors. Also the Sargan statistic suggests that there is no overidentification. This also holds if more control variables are included (see model specifications 2, 4 and 5 below).

Third, the Cragg-Donald Wald test indicates that the instrument is a strong instrument. Its F-statistic equals 8598.53 in the case of a dichotomous specification and 1.2E04 in case of the continuous specification.

In sum, the tests indicate that the instrument is a valid and strong instrument which can be applied in the IV analysis.

Results.

The results of the IV-analysis are presented in Table 3. The first two model specifications present the results for the dichotomous instrument, while the last three model specifications present the results for the continuous instrument. To capture the observed and unobserved heterogeneity, we include class fixed effects (models 2 and 4) and class and school fixed effects (model 5). The results are robust across all model specifications.

We observe that, instrumented for the participation in the CAI-tool, making more exercises lead to higher test scores. The coefficient is positive and significant for all specifications. As the average student makes 50 exercises (see Table 1), he/she increases the post-test scores by 0.035 (0.0007×50) which is about 3.5% as the post-test ranges between 0 and 1.

Table 3 further reveals that, as expected, students with higher abilities, i.e. a higher pre-test score, also have higher test scores. Controlled for class fixed effects, we observe that schools with higher national and school exam results have significantly higher test scores of students. Students in schools with higher material costs per student have lower test results, while schools with more teachers per students do not seem to have significantly higher test results. The IV-regression further reveals that, given the ability of the student, the percentage of students coming from disadvantaged neighbourhoods increases the post-test scores. The latter three observations are only significant in model specification 4. The remaining control variables are not significantly different from 0.

In sum, we find that, given the participation to the CAI-tool, making more exercises leads to higher test results. Working with a CAI-tool seems therefore effective.

Table 3: Instrumental variable analysis with class and school fixed effects.

Ind. Var: Post test scores	model1	model2	model3	model4	model5
Number of exercises	0.0007***	0.0007***	0.0008***	0.0006***	0.0006***
Pre-test – score (student level)	0.0049***	0.0027***	0.0051***	0.0028***	0.0027***
Pre-test - time (student level)		0.0000***		0.0000***	0.0000***
Average national exam (school level)		0.0701		0.1208*	(omitted)
Average school exam (school level)		0.247**		0.1197*	(omitted)
Number of teachers per student (fte - school level)		0.501		0.379	(omitted)
Costs for materials per student (school level)		-0.828***		-0.642***	(omitted)
Dropout percentage (school level)		0.0579**		-0.021	(omitted)
% disadvantaged students (school)		0.0016		0.0015**	(omitted)
% supportive ability track (school)		-0.0117		0.0008	(omitted)
Constant	0.3060***	-1.649**	0.2993***	-1.1262**	0.4284***

Instrument	Dummy (little - extensive)	Dummy (little - extensive)	Continuous	Continuous	Continuous
Fixed effects		Class		Class	Class and school
Number of observations	5478	5478	9898	9898	9898
R ² -adjusted	0.0817		0.0765		

Legend: * p<.1; ** p<.05; *** p<.0.01

5. Conclusion and policy recommendations

Whereas ICT infrastructure has improved considerably in secondary schools during the last decades, there still remains the enormous challenge for teachers and educational stakeholders to integrate this infrastructure in the teaching and learning activities. Undoubtedly, a key role in the integration of educational technology is played by the two stakeholders most involved in the education process: the teachers and the pupils. Overall, the literature shows that teachers play a critical role in determining (1) whether technology will be used in the classroom and (2) if so, how the educational technology will be exactly used.

As an empirical contribution, this paper exploited the variation in the use of an Computer-Assisted Instruction (CAI) tool. In particular, it examined the effectiveness of an addaptive Dutch CAI-tool for mathematics in lower secondary education (called Gotit?!). We observed that schools with, on average, lower attainments (as measured on nationwide standardized exams) rely more on the novel CAI-tool than schools with higher attainments. This suggests that schools see CAI-tools as a way to catch-up in learning outcomes. This finding is confirmed by the observation that schools with a higher share of students from disadvantaged neighborhoods (as defined by the central government) are more frequently working with the tool. Again, this suggests that schools are effectively using the tool to differentiate among students.

Moreover, this paper examined whether a higher exposure to the program leads to higher test outcomes, using an instrumental variable approach. We observe that, given the participation to the CAI-tool, making more exercises leads to higher test results. Working with a CAI-tool seems therefore effective.

Policy recommendations.

Given the important role of teachers, it is crucial that teachers dispose of adequate knowledge for using educational technology effectively. Therefore, it is important that policy makers as well as school directors invest both resources and time in the training of teachers. The results of this paper also suggest that policy makers can more actively encourage the use of ICT for schools with poor

learning outcomes or with a diverse student population. As the ICT hardware is nowadays available in most schools, ICT should be used in the most effective and efficient way. This paper shows that adaptive CAI-tools might be an effective tool to increase learning outcomes.

6. Literature

Angrist, J., & Lavy, V. (2002). New Evidence on Classroom Computers and Pupil Learning. *The Economic Journal*, 112(482), 735-765.

Beal, C.R., Arroyo, I.M., Cohen, P.R., & Woolf, B.P. (2010). Evaluation of AnimalWatch: An intelligent tutoring system for arithmetic and fractions. *Journal of Interactive Online Learning*, 9(1), 64-77.

Biagi, F., & Loi, M. (2013). Measuring ICT use and learning outcomes: Evidence from recent econometric studies. *European Journal of Education*, 48(1), 28-42.

Blok, H., Oostdam, R., Otter, M.E., & Overmaat, M. (2002). Computer-assisted instruction in support of beginning reading instruction: A review. *Review of Educational Research*, 72, 101-30.

Bouck, E. C., & Flanagan, S. M. (2009). Assistive technology and mathematics: What is there and where can we go in special education. *Journal of Special Education Technology*, 24(2), 17-30.

Campuzano, L., Dynarski, M., Agodini, R., & Rall, K., (2009), Effectiveness of reading and mathematics software products: Findings from two student cohorts. Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Christmann, E. P., & Badgett, J. L. (2000). The comparative effectiveness of CAI on collegiate academic performance. *Journal of Computing in Higher Education*, 11(2), 91–103.

Christmann, E. P., & Badgett, J. L. (2003). A meta-analytic comparison of the effects of computer-assisted instruction on elementary students' academic achievement. *Information Technology in Childhood Education Annual*, 2003(1), 91-104.

Cox, M. J. and Abbott, C. (Eds) (2004) , *ICT and Attainment – A Review of the Research Literature*. Coventry, Becta /London, DfES.

Cox, M.J., & Marshall, G. (2007). Effects of ICT: Do we know what we should know? , *Education and Information Technologies*, 12(2), 59-70.

De Witte, K., Geys, B., and Solondz, C. (2014). Public Expenditures, Educational Outcomes and Grade Inflation: Theory and Evidence from a Policy Intervention in the Netherlands. *Economics of Education Review*. In Press.

Doerr, H. M., & Zangor, R. (2000). Creating meaning for and with the graphing calculator. *Educational Studies in Mathematics*, 41(2), 143-163.

Dynarski, M., Agodini, R., Heaviside, S., Novak, T., Carey, N., Campuzano, L., Means, B., Murphy, R., Penuel, W., Javitz, H., Emery, D., Sussex, W. (2007), Effectiveness of reading and mathematics

software products: Findings from the first student cohort. Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Edwards, M. T., & Quesada, A. (2007). Dueling (Dualing) solids: enhancing student and teacher geometrical understanding with CABRI 3D. In Proceedings of the nineteenth annual International Conference on Technology in Collegiate Mathematics (ICTCM), February 15-18, 2007. Boston, MA.

European Commission (2000). Communication from the Commission-e-Learning - Designing tomorrow's education. COM(2000) 318 final.

Fuchs, T., & Wößmann, L. (2005). Computers and student learning: Bivariate and multivariate evidence on the availability and use of computers at home and at school. CESifo Working Paper 1321. Munich: CESifo.

Goldberg, A., Russell, M., & Cook, A. (2003). The effect of computers on student writing: A meta-analysis of studies from 1992-2002. *The Journal of Technology, Learning, and Assessment*, 2(1), 1-51.

Hennessy, S., Ruthven, K., & Brindley, S. (2005). Teacher perspectives on integrating ICT into subject teaching: Commitment, constraints, caution and change. *Journal of Curriculum Studies*, 37, 155-192.

Honey, M., Culp, K. M., & Carrigg, F. (2000). Perspectives on technology and education research: Lessons from the past and present. *Journal of Educational Computing Research*, 23(1), 5-14.

Kaput, J., Hegedus, S., & Lesh, R. (2007). Technology becoming infrastructural in mathematics education. In R. Lesh, E. Hamilton, & J. Kaput (Eds.), *Foundations for the future in mathematics education* (pp. 173–192). Mahwah, NJ: Lawrence Erlbaum Associates.

Kebritch, M., Hirumi, A., & Bai, H. (2010). The effects of modern mathematics computer games on mathematics achievement and class motivation. *Computers and Education*, 55(2), 427–443.

Koedinger, K.R., McLaughlin, E.A., & Heffernan, N.T. (2010). A quasi-experimental evaluation of an on-line formative assessment and tutoring system. *Journal of Educational Computing Research*, 43(4), 489-510.

Köklü, O., & Topçu, A. (2012). Effect of Cabri-assisted instruction on secondary school students' misconceptions about graphs of quadratic functions. *International Journal of Mathematical Education in Science and Technology*, 43(4), 1–13.

Kubiatko, M., & Vlckova, K. (2010). The relationship between ICT use and science knowledge for Czech students: A secondary analysis of PISA 2006. *International Journal of Science and Mathematics Education*, 8, 523-543.

Kulik, C. C., & Kulik, J. A. (1991). Effectiveness of computer-based instruction: an updated analysis. *Computers in Human Behavior*, 7, 75–94.

Kulik, J. A. (2003). Effects of using instructional technology in elementary and secondary schools: What controlled evaluation studies say. SRI International Report.

- Kulik, J.A. (1994). Meta-analytic studies of findings on computer-based instruction. In E.L. Baker and O'Neill, H.F. (Eds.), *Technology Assessment in Education and Training*, Hillsdale, NJ: Lawrence Erlbaum, 9-33.
- Leuven, E., Lindahl, M., Oosterbeek, H. & Webbink, D. (2004). The effect of extra funding for disadvantaged pupils on achievement. *The Review of Economics and Statistics*, 89, 721–736.
- Lopez-Morteo, G., & Lopez, G. (2007). Computer support for learning mathematics: a learning environment based on recreational learning objects. *Computers & Education*, 48(4), 618-641.
- Luu, K., & Freeman, J.G. (2011). An analysis of the relationship between information and communication technology (ICT) and scientific literacy in Canada and Austria. *Computers & Education*, 56, 1072-1082.
- Machin, S., McNally, S., & Silva, O. (2007). New technology in schools: Is there a payoff? *Economic Journal*, 117(522), 1145-1167.
- McAlister, M., Dunn, J., & Quinn, L. (2005). Student teachers' attitudes to and use of computers to teach mathematics in the primary classroom. *Technology, Pedagogy and Education*, 14(1), 77-105.
- Notten, N., & Kraaykamp, G. (2009). Home media and science performance: A cross national study. *Educational Research and Evaluation*, 15, 367-384.
- Pilli, O. & Aksu, M. (2013). The effects of computer-assisted instruction on the achievement, attitudes and retention of fourth grade mathematics students in North Cyprus. *Computers & Education*, 62, 62-71.
- Rayne, R., & Baggott, G. (2004). Computer-based and computer-assisted tests to assess procedural and conceptual knowledge.
- Reimer, K., & Mayer, P. S. (2005). Third-graders learn about fractions using virtual manipulatives: a classroom study [Electronic version]. *The Journal of Computers in Mathematics and Science Teaching*, 24(1), 5–25.
- Roschelle, J., Shechtman, N., Tatar, D., Hegedus, S., Hopkins, B., Empson, S., Knudsen, J., Gallagher, L. (2010). Integration of technology, curriculum, and professional development for advancing middle school mathematics: Three large-scale studies. *American Educational Research Journal*, 47(4), 833-878.
- Rouse, C. E., Krueger, A.B., & Markman, L. (2004). Putting computerized instruction to the test: A randomized evaluation of a 'scientifically-based' reading program. *Economics of Education Review*, 23(4), 323-338.
- Papastergiou, M. (2009). Digital Game-Based Learning in high school Computer Science education: Impact on educational effectiveness and student motivation. *Computers & Education*, 52(1), 1-12.
- Potocki, A., Ecalle, J., & Magnan, A. (2013). Effects of computer-assisted comprehension training in less skilled comprehenders in second grade: A one-year follow-up study. *Computers & Education*, 63, 131-140.

Schacter, J., & Fagnano, C. (1999). Does computer technology improve student learning and achievement? How, when, and under what conditions? *Journal of Educational Computing Research*, 20(4), 329-343.

Slavin, R.E., Chueng, A., Groff, C., & Lake C. (2008). Effective reading programs for middle and high schools: A best-evidence synthesis. *Reading Research Quarterly*, 43(3), 290-322.

Spiezia, V. (2010). Does computer use increase educational achievements? Student-level evidence from PISA. *OECD Journal: Economic Studies*, Volume 2010.

Wenglinsky, H. (1998). Does it compute? The relationship between educational technology and student achievement in mathematics. Princeton, NJ: Educational Testing Service.

Wittwer, J., & Senkbeil, M. (2008). Is students' computer use at home related to their mathematical performance at school? *Computers & Education*, 56, 736-746.